

caBIOperl Tutorial

- ▶ National Cancer Institute Center for Bioinformatics
- ▶ Gene Levinson, Ph.D., Contractor, Terrapin Systems

Goals and Objectives

- ▶ **In this tutorial, we will show you:**
 - what caBIOperl can do for your clinical/research program.
 - how caBIOperl fits into caCORE.
 - how to get started with caBIOperl.
 - how relational database tables are accessed with perl.
 - how to solve real research problems.

Contributors

- ▶ **Shan Jiang**
- ▶ **Sue Dubman**
- ▶ **Peter Covitz**
- ▶ **Ying Long**
- ▶ **Himanso Sahni**
- ▶ **Carl Schaefer**
- ▶ **caBIO Team, SAIC**
- ▶ **Systems and App Support Teams, Terrapin Systems**
- ▶ **General Contractor, Booz-Allen Hamilton**
- ▶ **NCI Center for Bioinformatics**
- ▶ **caBIO community**

Gene Levinson

- ▶ 15 years experience-- basic and clinical lab research.
- ▶ Ph.D. in Molecular Genetics from U.C. Irvine.
- ▶ Software dev., Technical training, IT experience
- ▶ Bioinformatics Application Scientist, Terrapin Systems– Contractor to NCICB.

Please, tell us what you would like to cover today!



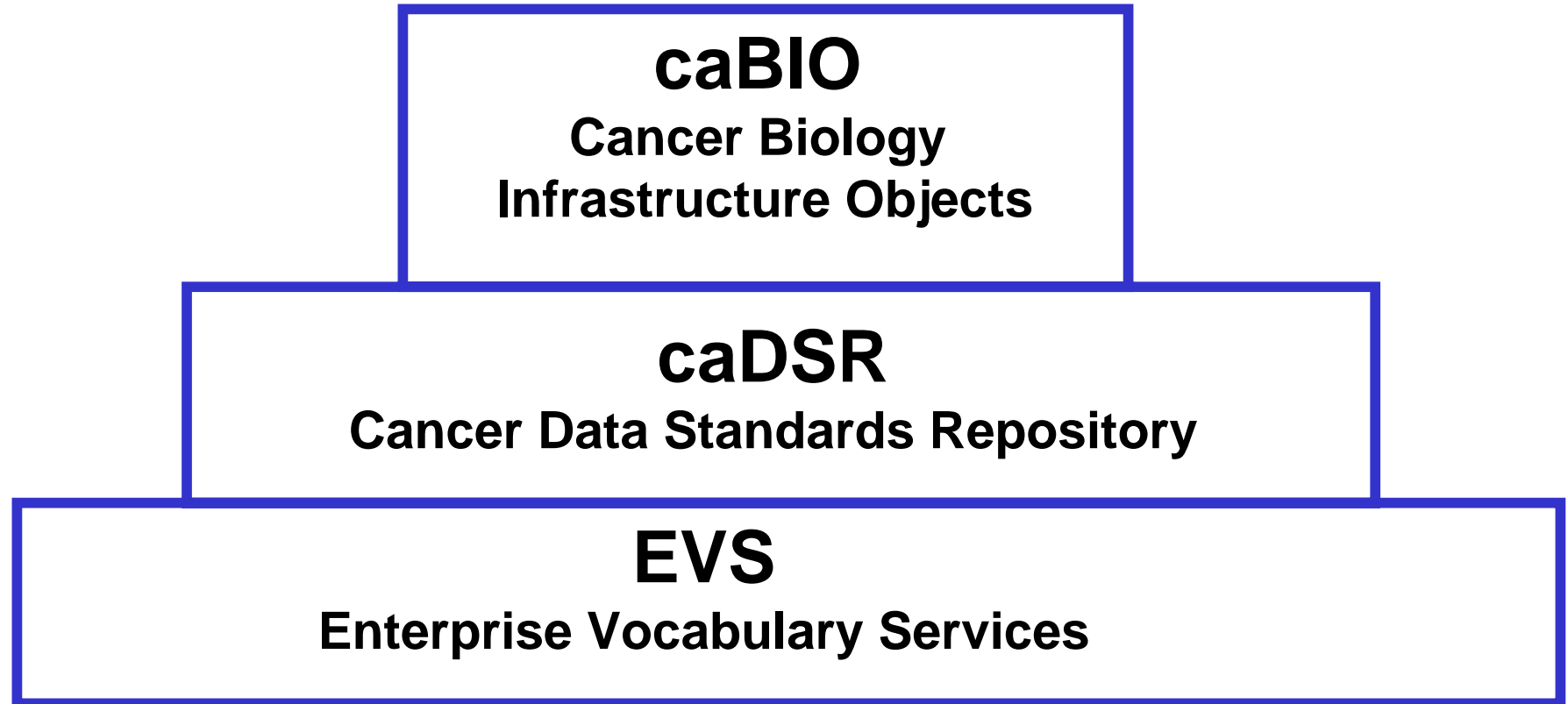


**Center
for
Bioinformatics**

Please— ask questions!!

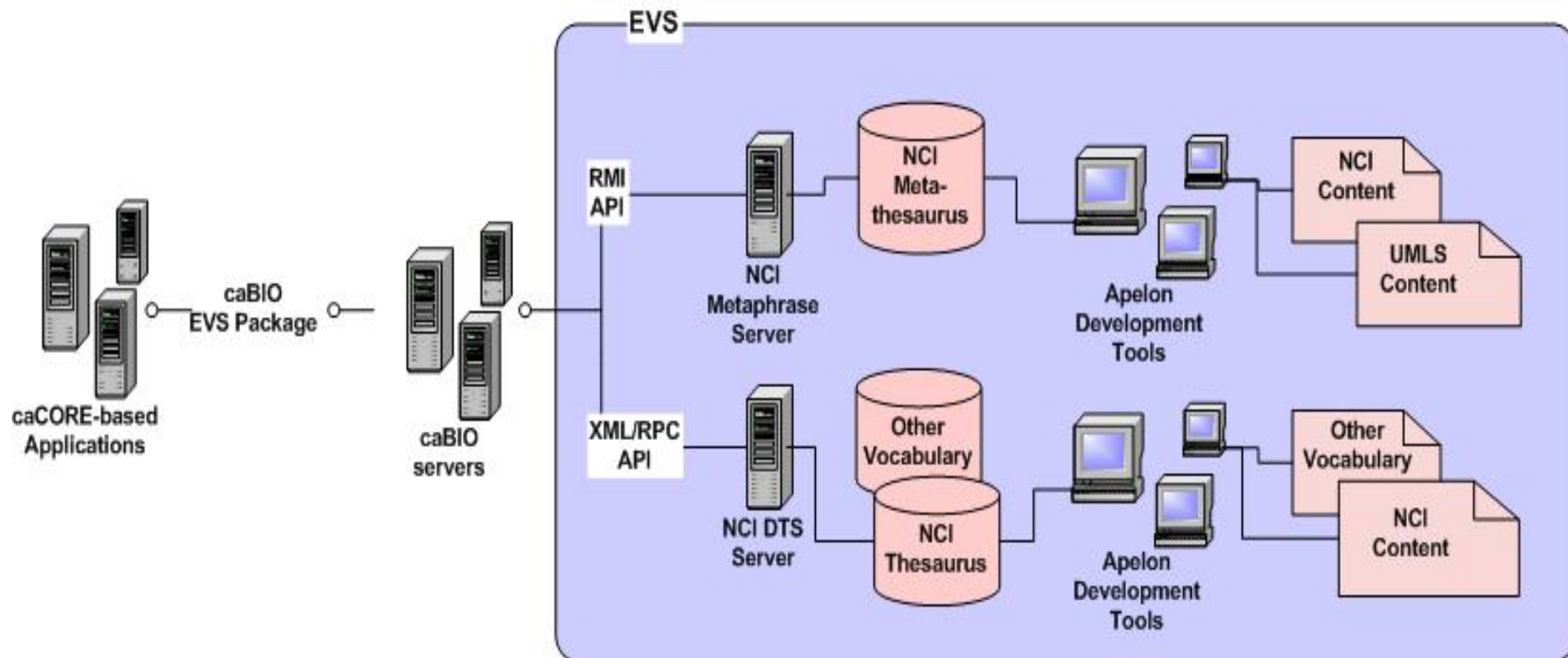


caBIO:
Rests on top of controlled vocabulary & standardized data elements



Enterprise Vocabulary Services

- ▶ <http://ncicb.nci.nih.gov/core/EVS>
- ▶ NCI thesaurus: Neoplasms, Drugs, Anatomy, Genes, Proteins, Techniques and administrative terminology, others
- ▶ Addresses need for a common vocabulary



Cancer Data Standards Repository

- ▶ metadata define data elements used in research studies
- ▶ metadata define questions being asked, and range of permissible answers.
- ▶ required for interpretation and computer processing of scientific data

The screenshot shows the CDE Browser web application running in a Microsoft Internet Explorer browser. The browser's address bar displays the URL <http://ncicb.nci.nih.gov/CDEBrowser/>. The application interface includes a sidebar on the left with a tree view of 'caDSR Contexts' containing various research programs and centers. The main content area features a 'Search Criteria' section with input fields for 'Valid Value', 'Public ID', 'Classification', and 'Context Use', along with a 'Search Field(s)' dropdown menu. A 'Data Element Search' button is located to the right of the search criteria. A 'Workflow Status' dropdown menu is also present. A red arrow points from the 'Search Criteria' section to a 'Results window' annotation. The 'Results window' annotation is a grey box with the text 'Results window' and a red arrow pointing to the right. The bottom of the browser window shows the status bar with the text 'User: Public User' and 'Build 2.0.1'. The bottom of the screen shows the Windows taskbar with various open applications and the system clock displaying '11:10 PM'.

Context “Drill Down”

Search Criteria

Results window

NATIONAL CANCER INSTITUTE

caDSR Contexts

- Assoc. American Cancer Institute
- NCI Center for Cancer Research (CCR)
- Cancer Imaging Program (CIP)
- City of Hope Med. Center (COH)
- NCI Cancer Therapy Evaluation Program (CTEP)
- Division of Cancer Epidemiology and Genetics (DCEG)
- NCI Division of Cancer Prevention (DCP)
- Early Detection Research Network (EDRN)
- Norris Cancer Center (NORRIS)
- Specialized Programs of Research Excellence (SPORes)
- test (TEST)
- Cancer Bioinformatics Infrastructure Objects (caBIO)

Workflow Status: ALL

Valid Value:

Public ID:

Classification:

Context Use: Owned & Used By

Search Field(s): ALL

Long Name

Preferred Name

Document Text

Search Data Elements

Clear

User: Public User

Build 2.0.1

National Institutes of Health (NIH)

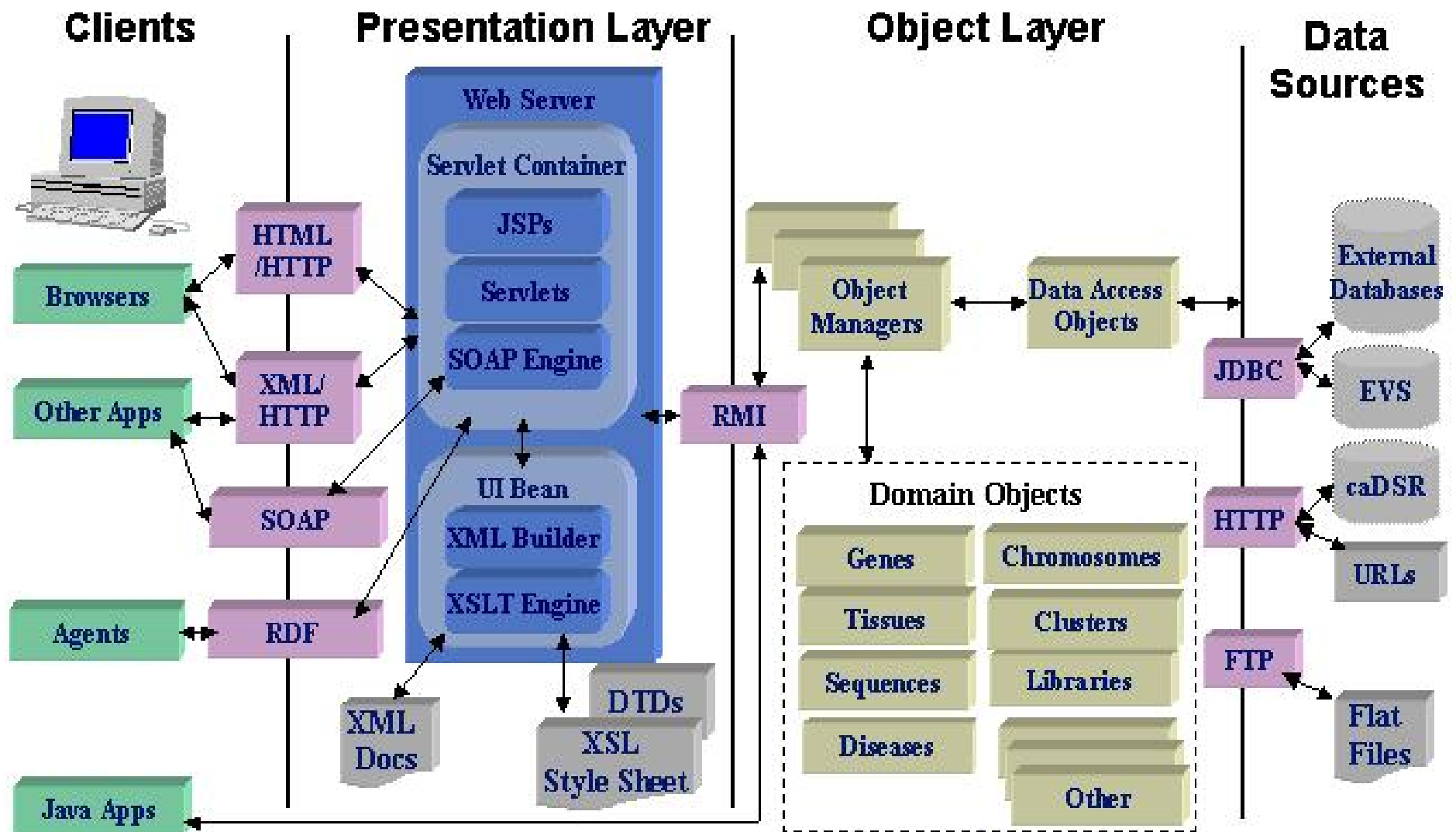
HEALTH & HUMAN SERVICES

FIRSTGOV

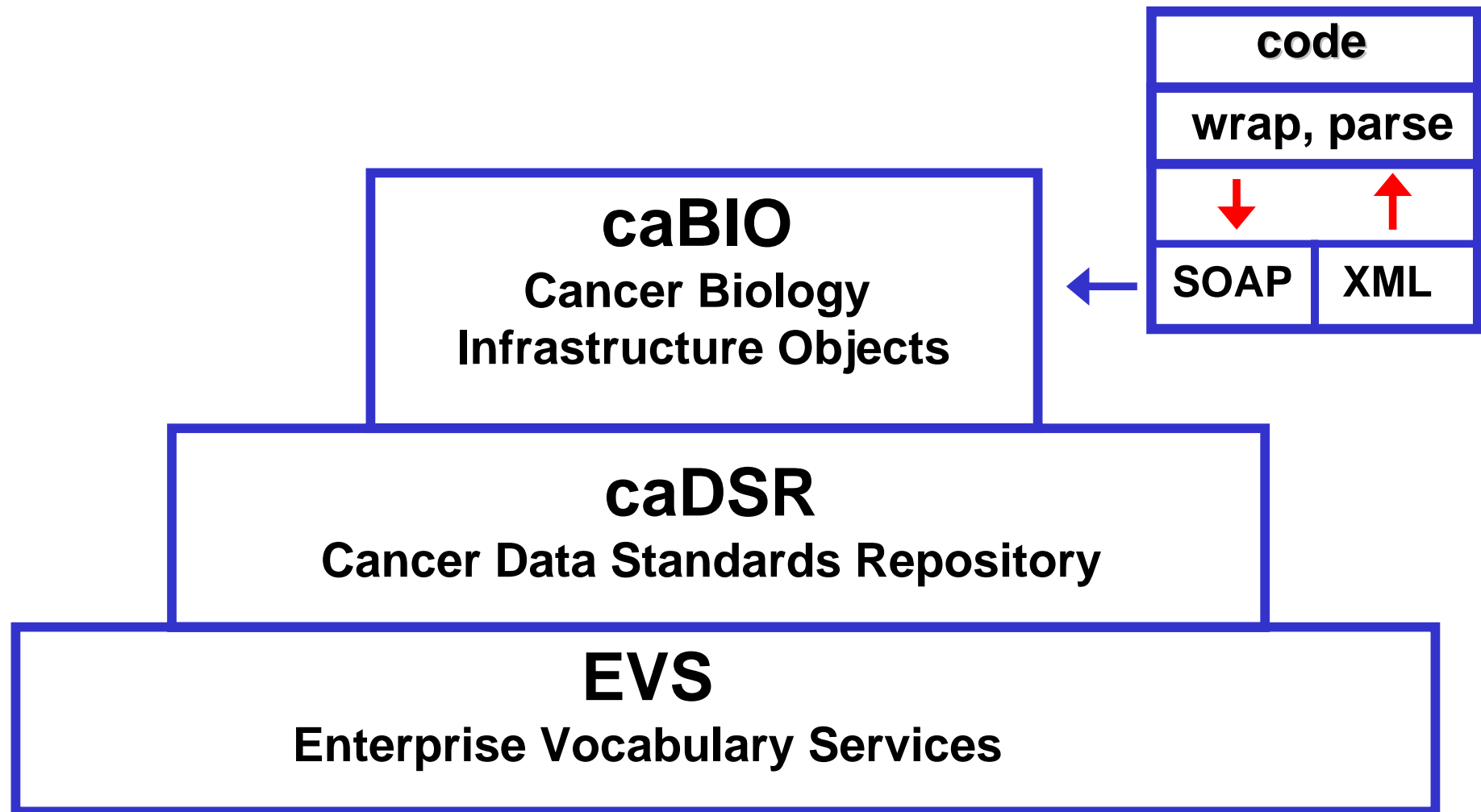
Your First Click to the U.S. Government

caBIG cancer Biomedical Informatics Grid

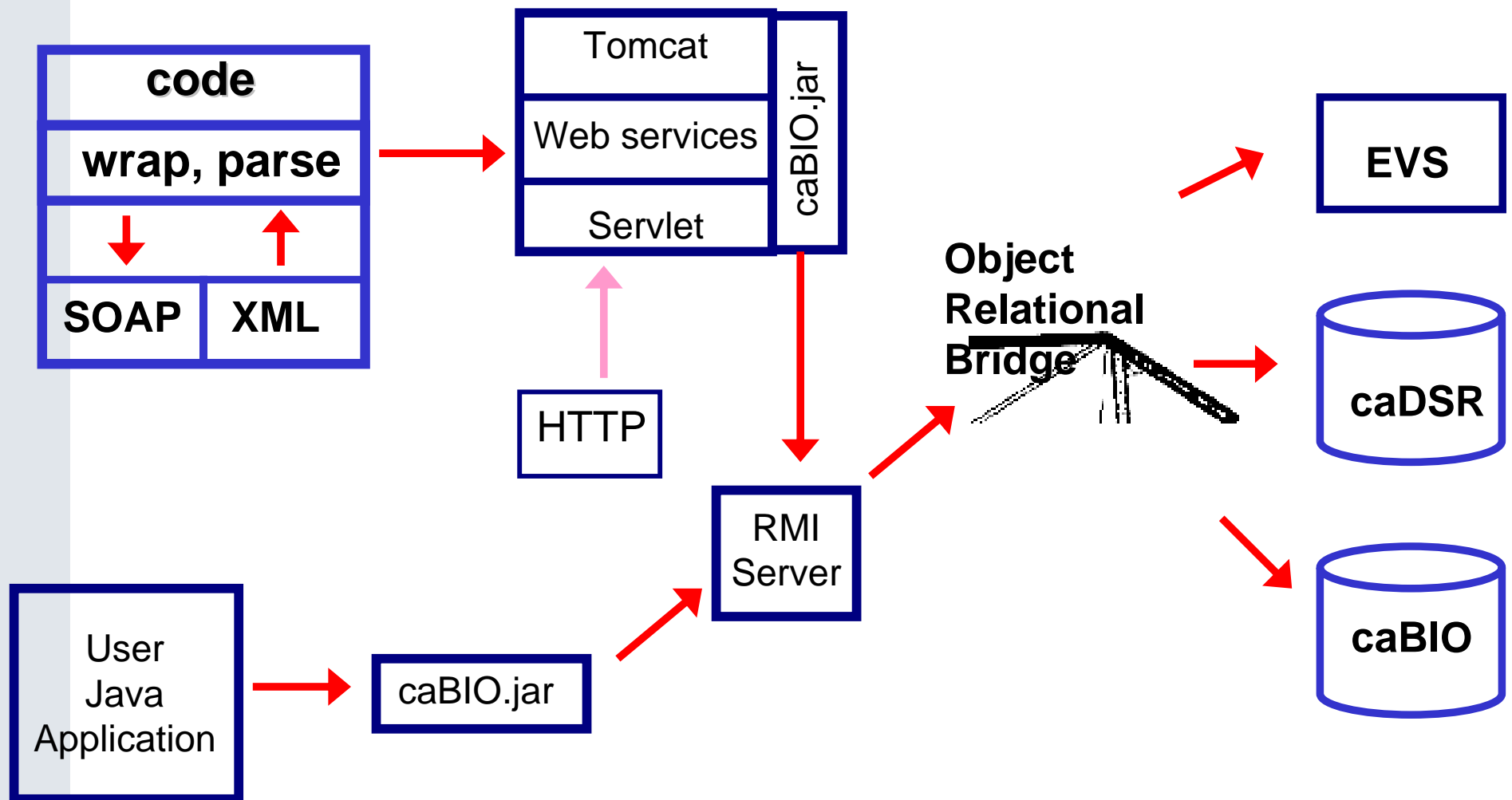
caBIO Architecture



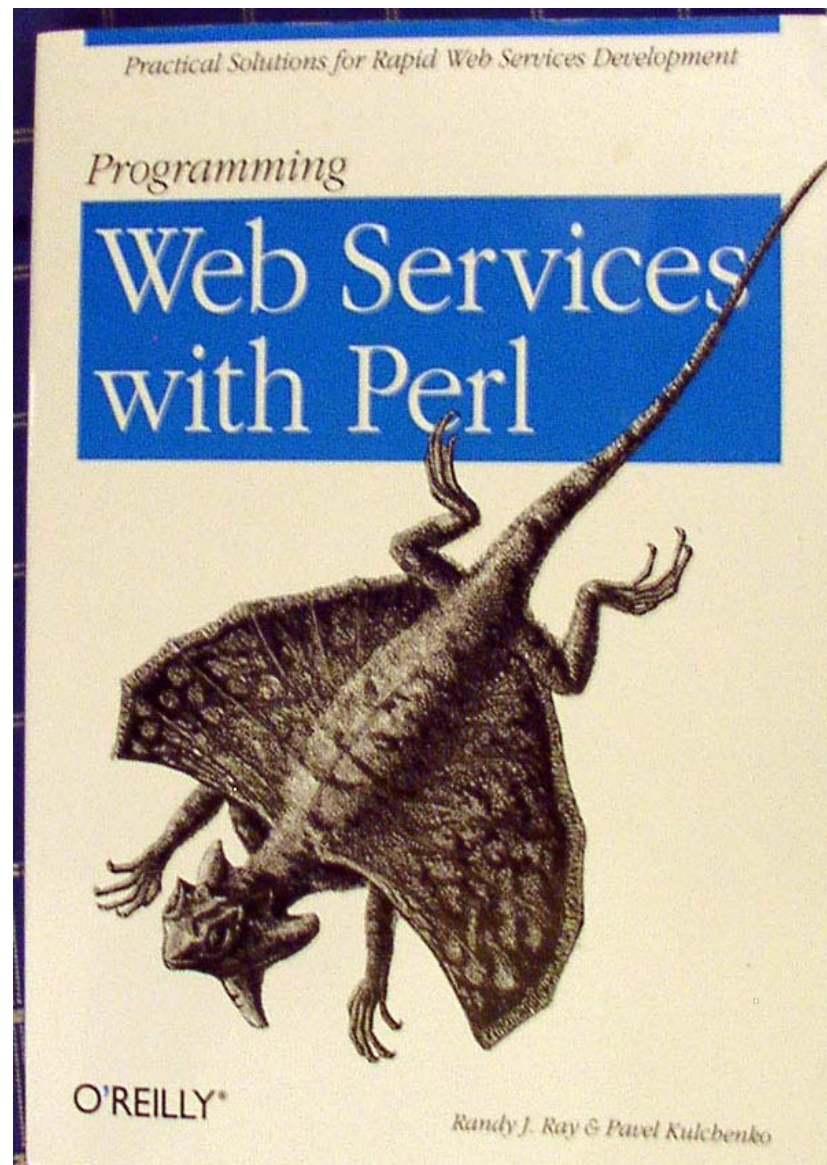
Access to Java elements for Perl programmers:



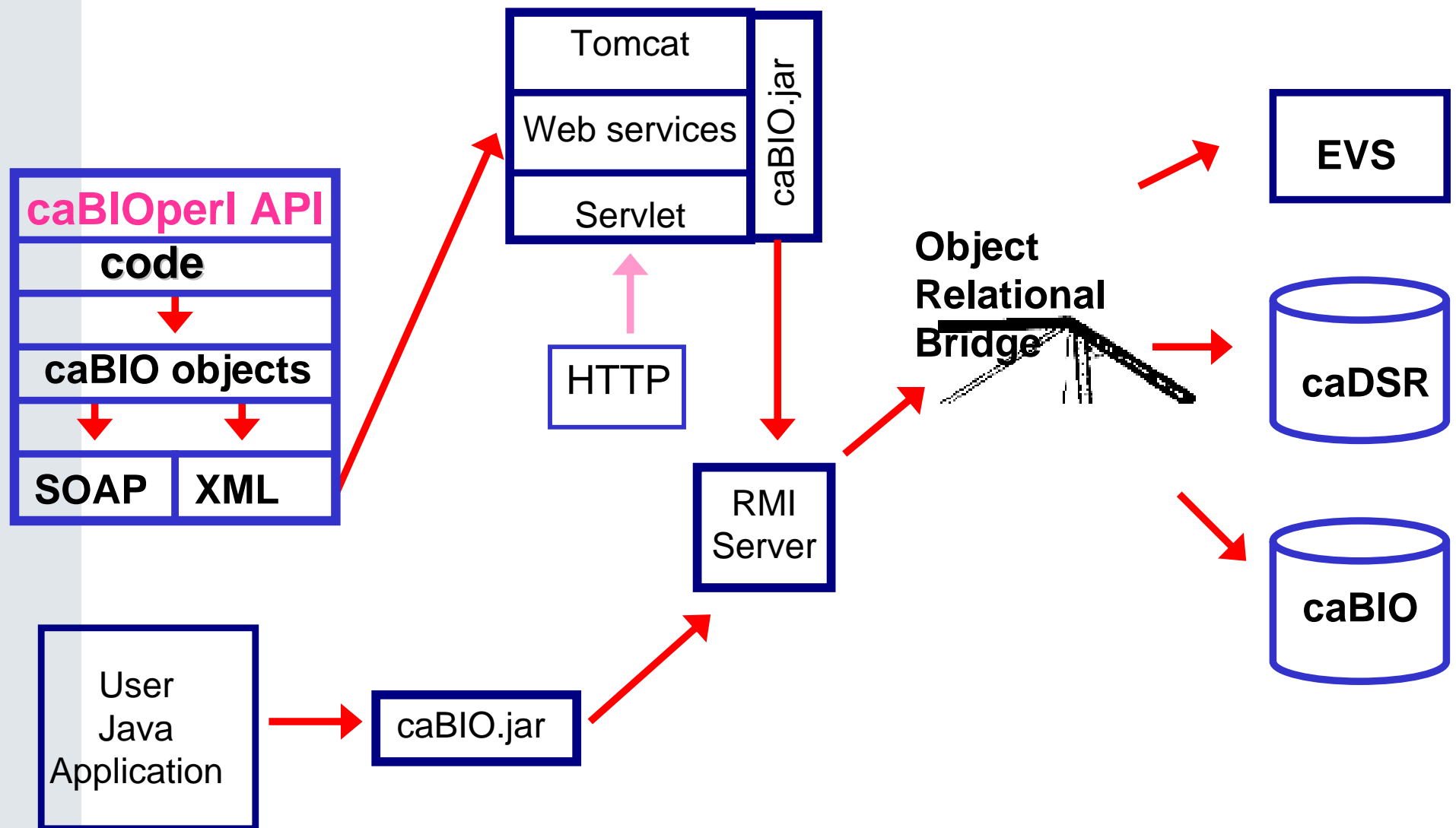
SOAP wrapper and XML access to caBIO via perl



Perl for Web Services can be a chore.



Native way to access caBIO via caBIOperl API



caBIOperl makes it easier to focus on investigator queries

gene object instantiated;
getPathways method called-
queries the database, returns
pathway name fields

```
@pathways = $gene ->getPathways() ;
```

Retrieving pathways:

Found caBIO Gene BRCA1

Found 5 associated Pathways

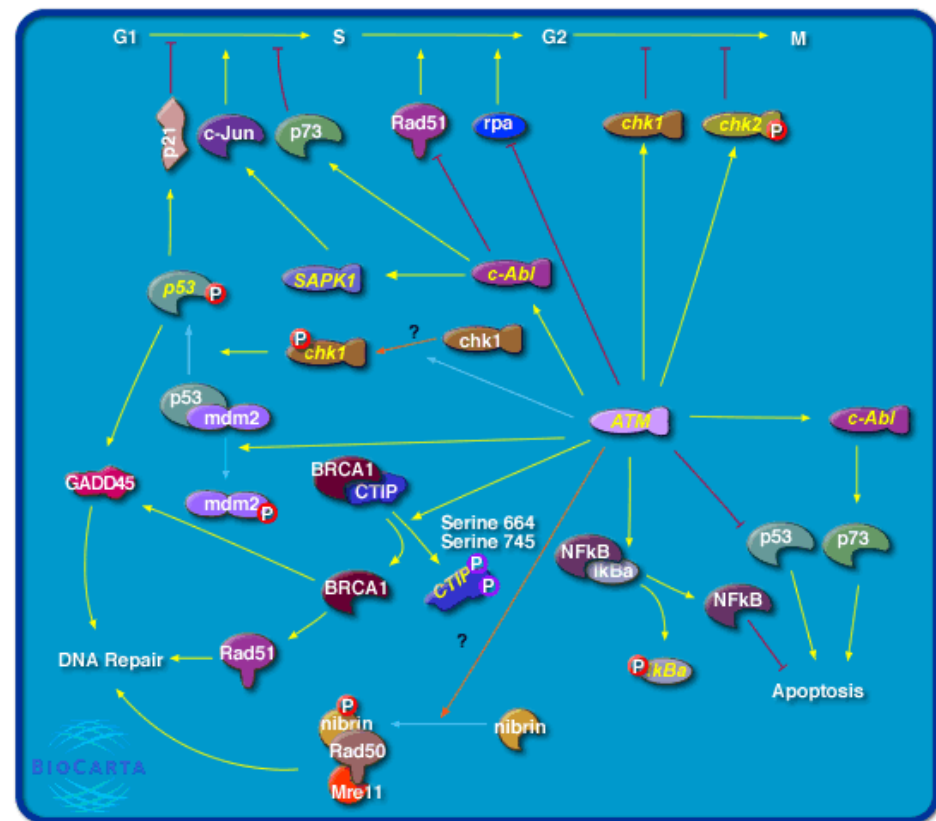
Name: h_atmPathway

Name: h_atrbrcaPathway

Name: h_bard1Pathway

Name: h_carm-erPathway

Name: h_g2Pathway

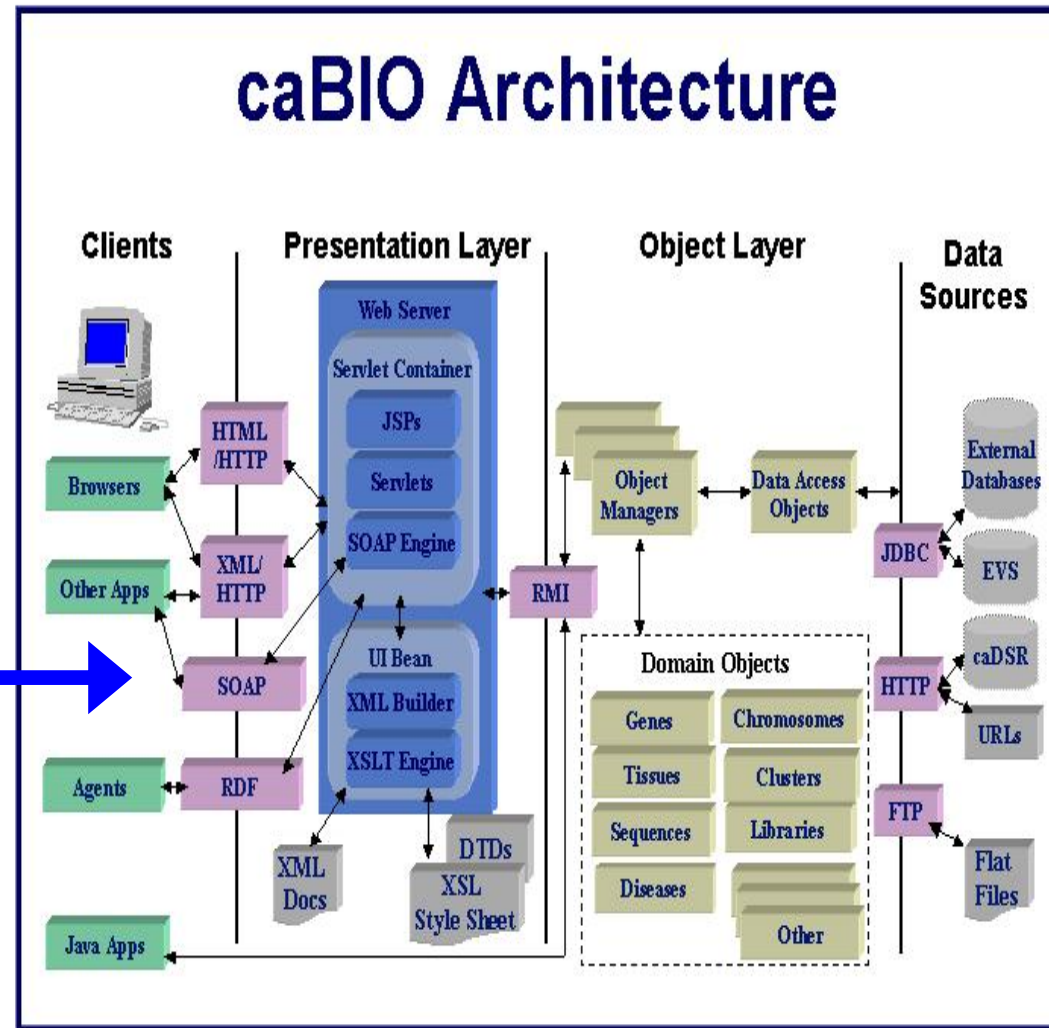


caBIOperl wraps lower-level SOAP API-- shields developers from SOAP calls and XML parsing .

XML and
SOAP
are “under the
hood”



caBIOperl



What can caBIOperl do for you?

- ▶ Use native perl code to access caBIO objects in an object-oriented fashion;
- ▶ Query a broad range of relational databases, using perl rather than SQL;
- ▶ Query large, remote databases including genes, pathways, clinical trials, etc;
- ▶ Navigate between objects, using perl code.

Getting Started: Set up perl

To see if Perl is present and accessible, type `perl -version` on command line.

Perl should be in your path.

Solaris/SunOS:

Binary at www.sunfreeware.com

Linux: probably already in package

If not, can get RPM at Linux-specific site, or at www.perl.org;

Windows: ActivePerl:

<http://www.activestate.com/Products/ActivePerl/>

Getting Started: Perl Modules

- ▶ Update your Perl installation with XML-DOM and SOAP-LITE:
 - Windows: at command prompt:
 - Type ppm, <ENTER>
 - Type search dom <ENTER>
 - Type install (number of XML-DOM entry) <ENTER>
 - Do the same search for SOAP; install SOAP-Lite.

Getting Started: Perl Modules

- ▶ Update your Perl installation with XML-DOM and SOAP-LITE:
 - Unix: at command prompt: universal method, any UNIX platform– at command prompt:

```
perl -MCPAN -e shell
```

Refer to the CPAN documentation for more details on how to use this command– at <http://www.perldoc.com/perl5.6/lib/CPAN.html>

Getting started: caBIOperl download

- ▶ <http://ncicb.nci.nih.gov/download/index.jsp>
- ▶ Unzip the package:
 - bin – **config scripts**
 - doc- **documentation**
 - examples
 - src
 - gov/nih/nci/caBIO/bean: Perl modules
 - t: perl test scripts for particular objects

Getting started: environment setup

- ▶ Suppose you unzipped the package to directory C:\caBIOperl. There are two ways to set up your environment in order to access the caBIOperl package:
 - Create or modify your environment variable PERL5LIB to include C:\caBIOperl\src.
 - Windows: Control Panel -> Systems -> Advanced; create/edit an environment variable PERL5LIB.
 - Modify the Perl environment variable @INC to include C:\caBIOperl\src

Getting started: caBIOperl config

- The standard distribution of the caBIOperl points to the caBIO server at cabio.nci.nih.gov, port 80. If you want to change them to point to your own designated location, you can run the `setServer.pl` and `setPort.pl` script in the `caBIOperl\bin` folder.

Example:

- `C:\caBIOperl\bin>setServer.pl
my_cabio_server.my_company.com`
- `C:\caBIOperl\bin>setPort.pl 8080`

caBIOperl API Documentation

- ▶ Documentation in HTML
- ▶ Readme describes setup and configuration
- ▶ Key aspects related to each of the objects are shown in the index:
 - Synopsis: minimal example of perl code to use that object
 - Object creation, attributes, associated methods described.
 - Associations indicate types of queries that can be imagined.
 - Other information such as author, license.

caBIOperl Documentation

[Agent](#)
[Anomaly](#)
[Chromosome](#)
[ClinicalTrialProtocol](#)
[Clone](#)
[CMAPOntology](#)
[CMAPOntologyRelationship](#)
[Disease](#)
[ExpressionFeature](#)
[ExpressionMeasurement](#)
[ExpressionMeasurementArray](#)
[Gene](#)
[GeneAlias](#)
[GeneHomolog](#)
[GoOntology](#)
[GoOntologyRelationship](#)
[Histopathology](#)
[Library](#)
[MapLocation](#)
[Organ](#)
[OrganRelationship](#)
[Pathway](#)
[Protein](#)
[ProteinHomolog](#)
[Protocol](#)
[ProtocolAssociation](#)
[Sequence](#)
[SNP](#)
[Target](#)
[Taxon](#)
[Tissue](#)
[Vocabulary](#)

- [Object creation and retrieval methods](#)
- [Attributes and related methods](#)
- [Associations and related methods](#)
- [EXPORT](#)
- [SEE ALSO](#)
- [AUTHOR](#)
- [COPYRIGHT AND LICENSE](#)

NAME

gov::nih::nci::caBIO::bean::Agent - Perl extension for Agent.

SYNOPSIS

```
use gov::nih::nci::caBIO::bean::Agent;  
use gov::nih::nci::caBIO::bean::AgentSearchCriteria;  
  
# create a AgentSearchCriteria instance to specify search elements  
my $search = new gov::nih::nci::caBIO::bean::AgentSearchCriteria;  
# specify a search element  
$search->setId(1080);  
# create a dummy bean instance to invoke its search methods  
my $bean = new gov::nih::nci::caBIO::bean::Agent;  
# return all beans that meets the search element  
my %beanlist = $bean->searchAgent($search);  
...
```

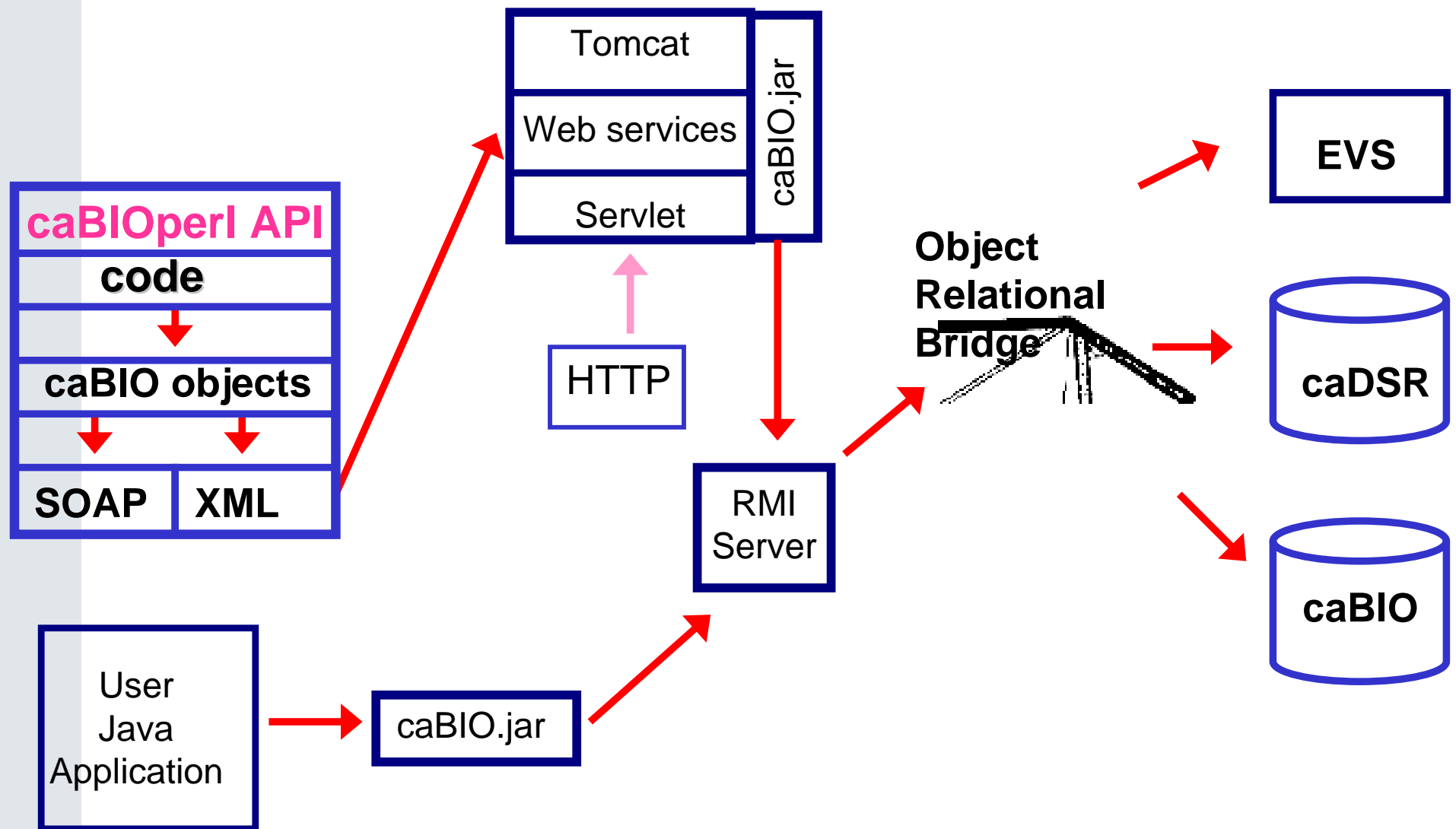


**Center
for
Bioinformatics**

Please— ask questions!!



Some examples of code, using the caBIOperl API



Gene symbols: used to search for pathways & taxons

```
#!/usr/local/bin/perl

# Map gene symbols to pathways using the caBioperl API.

use strict ;
use warnings ;
use gov::nih::nci::caBIO::bean::Gene ;
use gov::nih::nci::caBIO::bean::GeneSearchCriteria ;

my (@symbols, $taxonID, $sym, $faux, $search, $result, @genes, $gene, $taxon, @pathways, $pw) ;

# Search parameters
@symbols = qw(BRCA1 TP53) ;
$taxonID = 5 ;

foreach $sym (@symbols) {
    print "\nRetrieving pathways for gene '$sym'...\n" ;
    $faux = new gov::nih::nci::caBIO::bean::Gene() ;
    $search = new gov::nih::nci::caBIO::bean::GeneSearchCriteria() ;

    $search->setName($sym) ;
    if ( $result = $faux->search($search) ) {
        @genes = $result->getResultSet() ;
        foreach $gene (@genes) {
            printf " Found caBIO Gene %s\n", $gene->getName() ;
        }
    }
}
```

← Search object is
instantiated;

Gene symbols to pathways cont.

```
$taxon = $gene->getTaxon() ;  
if ($taxon) {  
    printf "    Found associated Taxon: %s\n",  
$taxon->getScientificName() ;  
    next unless $taxonID == $taxon->getId() ;  
} else {  
    printf "[WARNING] No associated Taxon found!\n",  
$gene->getName() ;  
    next ;  
}  
  
@pathways = $gene->getPathways() ;  
if (@pathways) {  
    printf "    Found %d associated Pathway(s):\n", scalar(@pathways) ;  
    foreach $pw (@pathways) {  
        printf "        Description: %s\n", $pw->getDisplayValue() ;  
$pw->getPathwayDescription() ;  
        printf "        SVG: %s\n", $pw->getName() ;  
    }  
} else {  
    print "    No associated Pathways found.\n" ;  
}  
}  
else {  
    print "[WARNING] No caBIO Gene found!\n" ;  
}  
}
```

← Taxon attribute is retrieved with getTaxon method;

← Pathways attribute retrieved with getPathways method;

Gene symbols to pathways: BRCA1 output

Retrieving pathways for gene 'BRCA1'...

Found caBIO Gene BRCA1

Found associated Taxon: Homo sapiens

Found 5 associated Pathway(s):

Description: ATM Signaling Pathway

SUG: h_atmPathway

Description: Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility

SUG: h_atrbrcaPathway

Description: BRCA1-dependent Ub-ligase activity

SUG: h_bard1Pathway

Description: CARM1 and Regulation of the Estrogen Receptor

SUG: h_carm-erPathway

Description: Cell Cycle: G2/M Checkpoint

SUG: h_g2Pathway

Found caBIO Gene Brca1

Found associated Taxon: Mus musculus

Gene symbols to pathways: TP53 output

```
Retrieving pathways for gene 'TP53'...
Found caBIO Gene TP53
Found associated Taxon: Homo sapiens
Found 18 associated Pathway(s):
  Description: Estrogen-responsive protein Efp controls cell
               cycle and breast tumors growth
  SUG: h_EfpPathway
  Description: Tumor Suppressor Arf Inhibits Ribosomal Biogenesis
  SUG: h_arfPathway
  Description: ATM Signaling Pathway
  SUG: h_atmPathway
  Description: Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility
  SUG: h_atrbrcaPathway
  Description: BTG family proteins and cell cycle regulation
  SUG: h_btg2Pathway
  Description: Apoptotic Signaling in Response to DNA Damage
  SUG: h_chemicalPathway
  Description: CTCF: First Multivalent Nuclear Factor
  SUG: h_ctcfPathway
  Description: Cell Cycle: G1/S Check Point
  SUG: h_g1Pathway
  Description: Cell Cycle: G2/M Checkpoint
  SUG: h_g2Pathway
  Description: p53 Signaling Pathway
  SUG: h_p53Pathway
  Description: Hypoxia and p53 in the Cardiovascular system
  SUG: h_p53hypoxiaPathway
  Description: Regulation of cell cycle progression by Plk3
  SUG: h_plk3Pathway
  Description: Regulation of transcriptional activity by PML
  SUG: h_pmlPathway
  Description: RB Tumor Suppressor/Checkpoint Signaling in
               response to DNA damage
  SUG: h_rbPathway
  Description: Double Stranded RNA Induced Gene Expression
  SUG: h_rnaPathway
  Description: Telomeres, Telomerase, Cellular Aging, and Immortality
  SUG: h_telPathway
  Description: Overview of telomerase protein component gene
               hTert Transcriptional Regulation
  SUG: h_tertpathway
  Description: Chaperones modulate interferon Signaling Pathway
  SUG: h_tidPathway
```


Several attribute searches using subroutines: gene = ATM

```
use gov::nih::nci::caBIO::bean::GeneSearchCriteria;

print "\nTest gov::nih::nci::caBIO::bean::Gene\n";
# main
## begin list test routines ##
my $USAGE =
"\nUsage: \n\nseveralGeneSearches.pl <gene symbol> (e.g. ATM) \n";

my $argc=@ARGV;

# one argument (gene symbol) is mandatory
if ($argc < 1)
{
    die $USAGE;
}
else
{
    $symbol = shift @ARGV;
}

if (getGenes() != 0 ) {print "\nFailed!\n";}
if (getGoOntologys() ) { print "\nFailed\n";}
if (getSNPs() ) { print "\nFailed\n";}
if (getMapLocations() ) { print "\nFailed\n";}
if (getProteins() ) { print "\nFailed\n";}
if (getGeneHomologs() ) { print "\nFailed\n";}
if (getGeneAliases() ) { print "\nFailed\n";}

print "\nTest complete\n";
exit 0;
```

ATM: getGenes subroutine

```
## begin search by attribute methods ##
# test search method
sub getGenes {
  # create an instance
  my $bean = new gov::nih::nci::caBIO::bean::Gene;

  # create a search criterion
  my $search = new gov::nih::nci::caBIO::bean::GeneSearchCriteria;
  $search->setName($symbol); ← 1. Set the symbol with which to search

  print "\n**** Test search ****\n";
  my $result = $bean->search($search); ← 5. Print the result

  # process result
  print "\n".$result->toString."\n\n";
  my @rs = $result->getResultSet; ← 2. Define an array to receive the result set
  my $bean2;
  foreach $bean2 (@rs) { ← 3. Set up a search object that uses that result set
    print "Name = ".$bean2->getName."\n";
    print "Title = ".$bean2->getTitle."\n";
    print "LocusLinkId = ".$bean2->getLocusLinkId."\n";
    print "ClusterId = ".$bean2->getClusterId."\n\n";
  }
  print "\nSucceed\n"; ← 4. Call each attribute to search for as a method
  return 0;
}
```


ATM: getGenes subroutine: retrieved name, title, locus link ID, cluster ID

```
Test gov::nih::nci::caBIO::bean::Gene
```

```
***** Test search *****
```

```
SearchResult common attributes:
```

```
startsAt=0      endsAt=2      hasMore=false      num. of obj=2
```

```
Name = ATM
```

```
Title = ataxia telangiectasia mutated (includes complementation  
groups A, C and D)
```

```
LocusLinkId = 472
```

```
ClusterId = 526394
```

```
Name = Atm
```

```
Title = ataxia telangiectasia mutated homolog (human)
```

```
LocusLinkId = 11920
```

```
ClusterId = 5088
```

```
Succeed
```

ATM: getGoOntologys subroutine

```
# test getGoOntologys method
sub getGoOntologys {
    # create an instance
    my $bean = new gov::nih::nci::caBIO::bean::Gene;

    print "\n**** Test getGoOntologys ****\n";
    $bean->setName($symbol);
    my @beans = $bean->getGoOntologys;
    if ( @beans ) {
        print "\nGot a total of " . ($#beans+1) . " GoOntologys.\n\n";
        my $bn;
        foreach $bn (@beans) {
            print "MouseGenesCount = " . $bn->getMouseGenesCount . "\n";
            print "HomoSapienGenesCount = " . $bn->getHomoSapienGenesCount . "\n";
            print "Name = " . $bn->getName . "\n\n";
        }
    } else {
        print "\nFound no GoOntologys.\n";
    }
    print "\nSucceed\n";
    return 0;
}
```

ATM: getGoOntologys subroutine:

```
**** Test getGoOntologys ****  
Got a total of 15 GoOntologys.  
MouseGenesCount = 98  
HomoSapienGenesCount = 219  
Name = regulation of cell cycle  
  
MouseGenesCount = 1281  
HomoSapienGenesCount = 1623  
Name = DNA binding  
  
MouseGenesCount = 15  
HomoSapienGenesCount = 22  
Name = inositol/phosphatidylinositol kinase activity  
  
MouseGenesCount = 457  
HomoSapienGenesCount = 485  
Name = protein kinase activity  
  
MouseGenesCount = 331  
HomoSapienGenesCount = 365  
Name = protein serine/threonine kinase activity  
  
MouseGenesCount = 2424  
HomoSapienGenesCount = 2948  
Name = nucleus  
  
MouseGenesCount = 97  
HomoSapienGenesCount = 156  
Name = DNA repair  
  
MouseGenesCount = 15  
HomoSapienGenesCount = 25  
Name = response to DNA damage stimulus
```

ATM: getSNPs subroutine

```
# test getSNPs method
sub getSNPs {
    # create an instance
    my $bean = new gov::nih::nci::caBIO::bean::Gene;

    print "\n**** Test getSNPs ****\n";
    $bean->setName($symbol);
    my @beans = $bean->getSNPs;
    if ( @beans ) {
        print "\nGot a total of " . ($#beans+1) . " SNPs.\n";
        my $bn;
        foreach $bn (@beans) {
            print "Base1 = " . $bn->getBase1 . "\n";
            print "Base2 = " . $bn->getBase2 . "\n";
            print "Offset = " . $bn->getOffset . "\n";
            print "Score = " . $bn->getScore . "\n\n";
        }
    } else {
        print "\nFound no SNPs.\n";
    }
    print "\nSucceed\n";
    return 0;
}
```

ATM: getSNPs subroutine

```
***** Test getSNPs *****  
  
Got a total of 3 SNPs.  
Base1 = G  
Base2 = T  
Offset = 13020  
Score = 1  
  
Base1 = C  
Base2 = T  
Offset = 12670  
Score = 0  
  
Base1 = G  
Base2 = A  
Offset = 13032  
Score = 0  
  
Succeed
```

ATM: getMapLocations subroutine

```
# test getMapLocations method
sub getMapLocations {
    # create an instance
    my $bean = new gov::nih::nci::caBIO::bean::Gene;

    print "\n**** Test getMapLocations ****\n";
    $bean->setName($symbol);
    my @beans = $bean->getMapLocations;
    if ( @beans ) {
        print "\nGot a total of " . ($#beans+1) . " MapLocations.\n\n";
        my $bn;
        foreach $bn (@beans) {
            print "Location = " . $bn->getLocation . "\n";
            print "Type = " . $bn->getType . "\n\n";
        }
    } else {
        print "\nFound no MapLocations.\n";
    }
    print "\nSucceed\n";
    return 0;
}
```


ATM: getProteins subroutine

```
# test getProteins method
sub getProteins {
    # create an instance
    my $bean = new gov::nih::nci::caBIO::bean::Gene;

    print "\n**** Test getProteins ****\n";
    $bean->setName($symbol);
    my @beans = $bean->getProteins;
    if ( @beans ) {
        print "\nGot a total of " . ($#beans+1) . " Proteins.\n";
        my $bn;
        foreach $bn (@beans) {
            print "ProteinInfoId = " . $bn->getProteinInfoId . "\n";
            print "GeneInfoId = " . $bn->getGeneInfoId . "\n\n";
        }
    } else {
        print "\nFound no Proteins.\n";
    }
    print "\nSucceed\n";
    return 0;
}
```

ATM: getMapLocations and getProteins subroutines

```
**** Test getMapLocations ****  
  
Got a total of 1 MapLocations.  
Location = 11q22-q23  
Type = C  
  
Succeed  
  
**** Test getProteins ****  
  
Got a total of 2 Proteins.  
ProteinInfoId = 13878337  
GeneInfoId = sp:Q13315  
  
ProteinInfoId = 13878338  
GeneInfoId = sp:Q62388  
  
Succeed
```


ATM: getGeneHomologs subroutine

```
# test getGeneHomologs method
sub getGeneHomologs {
    # create an instance
    my $bean = new gov::nih::nci::caBIO::bean::Gene;

    print "\n**** Test getGeneHomologs ****\n";
    $bean->setName($symbol);
    my @beans = $bean->getGeneHomologs;
    if ( @beans ) {
        print "\nGot a total of " . ($#beans+1) . " GeneHomologs.\n";
        my $bn;
        foreach $bn (@beans) {
            print "SimilarityPercentage = " . $bn->getSimilarityPercentage . "\t";
        }
    } else {
        print "\nFound no GeneHomologs.\n";
    }
    print "\nSucceed\n";
    return 0;
}
```

ATM: getGeneHomologs subroutine

```
**** Test getGeneHomologs ****
```

```
Got a total of 2 GeneHomologs.
```

```
SimilarityPercentage = 84.31
```

```
Succeed
```

```
SimilarityPercentage = 84.31
```

ATM: getAliases subroutine

```
# test getGeneAliases method
sub getGeneAliases {
    # create an instance
    my $bean = new gov::nih::nci::caBIO::bean::Gene;

    print "\n**** Test getGeneAliases ****\n";
    $bean->setName($symbol);
    my @beans = $bean->getGeneAliases;
    if ( @beans ) {
        print "\nGot a total of " . ($#beans+1) . " GeneAliases.\n";
        my $bn;
        foreach $bn (@beans) {
            print "Type = " . $bn->getType . "\n";
            print "Description = " . $bn->getDescription . "\n";
            print "Name = " . $bn->getName . "\n\n";
        }
    } else {
        print "\nFound no GeneAliases.\n";
    }
    print "\nSucceed\n";
    return 0;
}
```

ATM: getAliases subroutine

```
**** Test getGeneAliases ****  
  
Got a total of 4 GeneAliases.  
Type = BioCarta  
Description = Mm.atm  
Name = Mm.atm  
  
Type = BioCarta  
Description = atm  
Name = atm  
  
Type = CGAP  
Description = ATM  
Name = ATM  
  
Type = CGAP  
Description = Atm  
Name = Atm  
  
Succeed  
  
Test complete
```



**Center
for
Bioinformatics**

Please— ask questions!!



NCICB App Support: tel. 301-451-4384

email ncicb@pop.nci.nih.gov

Address <http://ncicbsupport.nci.nih.gov/sw/>

Google nci application support Search Web PageRank Site popups allowed AutoFill Options nci application support

cancer.gov

NATIONAL CANCER INSTITUTE

Center for Bioinformatics

Home Organization Initiatives Infrastructure Support Download

NCICB Support Home:

Application Support for:

- Cancer Centralized Clinical Database (C3D)
 - Oracle Clinical RDC
 - Oracle Discover
- caArray
 - caVWorkbench
 - Gene Expression Data Portal (GEDP)
- caDSR Applications
 - caDSR Administration Tool
 - CDE Browser
 - CDE Compliance Review Response Tool
 - CDE Curation Tool
- Cancer Images Database (caIMAGE)
- Cancer Laboratory Information Management System (caLIMS)
- Cancer Model Organisms Database (caMOD)
- Cancer Molecular Analysis Project (CMAP)
- DCP Enterprise System Knowledgebase (DESK)
- Enterprise Vocabulary Services (EVS)
 - NCI Metathesaurus
 - NCI Thesaurus

Programming & API Support for:

- Cancer Bioinformatics Infrastructure Objects (caBIO)
- Cancer Data Standards Repository (caDSR)
- Enterprise Vocabulary Services (EVS)

[Return to NCICB Home](#)

Welcome to the NCICB Application Support Site

The NCICB Application Support Group is dedicated to helping you use NCICB tools and resources to maximum benefit. We're here to help you find what you need, use and re-use our software, integrate data and applications, and facilitate the sharing of information and resources.

Telephone: 301-451-4384 or toll free: 888-478-4423

Telephone support is available Monday to Friday, 8 am – 8 pm Eastern Time, excluding government holidays. You may leave a message, send an email or submit a support request via the Web at any time.

Email: ncicb@pop.nci.nih.gov

When submitting support requests via email, please include:

- your contact information
- the name of the application/tool you are using
- the URL if it is a Web-based application
- a description of the problem and steps to recreate it
- the text of any error messages you have received

Web Self-Service:

[Search FAQs](#) Click [HERE](#) to search our dynamic knowledgebase of Frequently Asked Questions (FAQs).

[Training Docs](#) Click [HERE](#) to view recent training presentations.

NCICB User Applications Manual

The NCICB User Applications Manual provides a thorough introduction to NCICB applications. Click [HERE](#) to view it in **PDF format**.

caBIOperl uses Apache OJB to query an Oracle relational database – usually a large, remote database.

GENE

gene_id	gene_symbol	chromosome_id
41645	tp53	19

CHROMOSOME

chromosome_id	taxon_id	chromosome_number
19	5	17

Object relational bridge

```
<class-descriptor
  class="gov.nih.nci.caBIO.bean.Gene"
  table="GENE"
  initialization-method="initDbCrossRefs">
<!--
  <extent-class class-ref="gov.nih.nci.caBIO.bean.GeneHomolog" />
-->
  <field-descriptor
    name="id"
    column="GENE_ID"
    jdbc-type="BIGINT"
    primarykey="true"/>

    <field-descriptor
      name="locusLinkId"
      column="LOCUS_LINK_ID"
      jdbc-type="VARCHAR"/>

    <field-descriptor
      name="OMIMId"
      column="OMIM_ID"
      jdbc-type="VARCHAR"/>

    <field-descriptor
      name="title"
      column="GENE_TITLE"
      jdbc-type="VARCHAR"/>

    <field-descriptor
      name="name"
      column="GENE_SYMBOL"
      jdbc-type="VARCHAR"/>
```


Summary

- ▶ caBIOperl API allows access to WebServices without the need for SOAP wrappers.
- ▶ caCORE provides object relational bridge such that searches can be performed with object instantiation.
- ▶ Oracle relational database records are represented by Java objects– can be queried with Perl.
- ▶ A range of objects of interest to cancer research-- and relationships between those objects– can be queried.
- ▶ App support available 8 AM- 8 PM Eastern time, Monday – Friday.






**Center
for
Bioinformatics**

Please— ask questions!!



An optional review of the caBIOperl objects will follow.



**Center
for
Bioinformatics**

[Home](#) [Organization](#) [Initiatives](#) [Infrastructure](#) [Support](#) [Download](#)

NCICB :

News


- [NCICB User Applications Manual](#)
- [caBIOperl released](#)
- [caBIG Pilot Structure](#)
- [2003 Jamboree Presentations](#)
- [NCICB Seminars](#)
- [Publications](#)

Quick Links


- [NCICB Support](#)
- [cancer Biomedical Informatics Grid \(caBIG\)](#)
- [Cancer Genome Anatomy Project](#)
- [Cancer Molecular Analysis Project](#)
- [Cancer Models Database](#)
- [Gene Expression Data Portal](#)
- [BIOgopher](#)
- [caBIO Quick Start](#)

The NCI Center for Bioinformatics


The NCI Center for Bioinformatics (NCICB) provides biomedical informatics support and integration capabilities to the cancer research community. We work with both intramural and extramural groups to develop Initiative-Specific Modules. These modules are connected through intelligent interfaces, coordinated through an NCI Core Module and deployed through open source tools and systems. The NCICB also serves as a focal point for cancer research informatics planning worldwide. We work with research organizations, biomedical informatics groups and standards bodies to facilitate the identification and adoption of information exchange standards, thus connecting research information sources wherever they may reside.



Organization
Overview of the NCICB's mission, organization, governance, and strategies for executing its charge.



Supported Research Initiatives
Access to the WWW sites and related data, tools, and infrastructure developed by the NCICB to support the collection of NCI research initiatives.



Infrastructure
[caCORE](#) is the infrastructure backbone supporting data management and application development at NCICB. caCORE includes vocabulary services ([EVS](#)), metadata management ([caDSR](#)), and biomedical data "objects" ([caBIO](#)) implemented in a robust [enterprise software architecture](#). The [caCORE Technical Guide](#) provides detailed documentation of the architecture and [APIs](#). All NCICB-developed caCORE components are [distributed](#) under open-source licenses that support unrestricted usage by both non-profit and commercial entities.

Thank you for using caBIO!!

Appendix

- ▶ A survey of some caBIO objects used by caBIOperl

caBIOperl objects

- ▶ Agent
- ▶ Anomaly
- ▶ Chromosome
- ▶ ClinicalTrialProtocol
- ▶ Clone
- ▶ CMAPOntology
- ▶ Disease
- ▶ ExpressionFeature
- ▶ ExpressionMeasurement
- ▶ ExpressionMeasurementArray
- ▶ Gene
- ▶ GeneAlias
- ▶ GeneHomolog
- ▶ GoOntology
- ▶ Histopathology
- ▶ Library
- ▶ MapLocation
- ▶ Organ
- ▶ Pathway
- ▶ Protein
- ▶ ProteinHomolog
- ▶ Protocol
- ▶ ProtocolAssociation
- ▶ Sequence
- ▶ SNP
- ▶ Target
- ▶ Taxon
- ▶ Tissue
- ▶ Vocabulary

Agent

- ▶ A therapeutic agent (drug, intervention therapy) used in a clinical trial protocol.
- ▶ Application: used primarily by CMAP and EVS applications.
- ▶ Related domain objects: ClinicalTrialProtocol, Target



CABIOOP:agent (oracle table)

A...	A...	AGENT_NAME	AGENT_SOURCE	AGENT_COMMENT	C...	NSC_NUMBER	EVS_ID
2			One study in NSCLC has rep		1		
3		2-Methoxy Estradiol	EntreMed	Inhibition of cellular proliferation, inhibit tubulin function, up-regulate	1	659853	C965
4		2B1	NCI	bispecific murine mAb	1		C2399
5		2C4	Genentech	Inhibits HER receptor dimerization	1		C2193
6		5-Aza-2'-deoxycytidine(Decitabi		DNA demethylating/hypomethylating agent	1	127716	C981
8		776C85 (eniluracil)	Glaxo Wellcome	Phase II in refractory breast cancer	1		C1773
9		9-cis retinoic acid (Panretin)	NCI	Phase I/II study	1	659772	C1574
10		ABT-627	Abbott Laboratories	Endothelin A receptor antagonist	1	720763	C1779
11		ABX-EGF	Abgenix/Immunex	MAb (human)	1		C1857
12		APO2L/TRAIL	Genentech/Immunex	Apoptosis inducing ligand 2 (also called TRAIL); member of TNF	1		C1685
15		Anastrozole (Arimidex)	Zeneca Pharmaceuticals	Phase III; first line and adjuvant therapy	1		C1607
16		Angiostatin	EntreMed	Inhibition of endothelial cell function	1		C17910
17		Angiozyme	Ribozyme Pharmaceuticals	Ribozyme targeting Flt-1 mRNA	1		C1865
18		Anti-idiotypic MUC-1		Ken Foon	1		C2709

Anomaly

- ▶ An irregularity in either the expression of a gene or its structure (i.e., a mutation).
- ▶ Application: defined and used by the CMAP project.
- ▶ Related domain objects: Histopathology, Target

CABIOOP:anomaly

	ANOMALY_ID	TARGET_ID	ANOMALY_TYPE	CONTEXT_CODE
	1	1	Amplification	1
	2	2	Reduced Expression	1
	3	3	Deleted	2
	4	4	Deleted	2
	5	5	Overexpression	2
	6	6		1
	7	7	Mutation	1
	8	8	Amplification/Mutation	2
	9	8	Expression	1
	10	8	Overexpression	1
	11	9		1
	12	10		1
	13	11		1
	14	12		1
	15	13	Overexpression	2
	16	14	Reduced Expression	1
	17	15	Amplified	2
	18	16		1
	19	17	Activated	2
	20	18		1

Chromosome

- ▶ An object representing a specific chromosome for a specific taxon; provides access to all known genes contained in the chromosome and to the taxon.
- ▶ Application: used by CMAP and other applications to reason about the molecular basis of cancer.
- ▶ Related domain objects: Gene, Taxon

CABLOOP:chromosome

CHROMOSOME_ID	TAXON_ID	CHROMOSOME_NUMBER
0	5	8
1	5	4
2	5	19
3	5	12
4	5	14
5	5	18
6	5	9
7	5	11
8	5	1
9	5	6
10	5	7
11	5	X
12	5	2
13	5	21
14	5	3
15	5	20
16	5	16
17	5	13
18	5	22
19	5	17
20	5	15
21	5	10
22	5	5
23	5	Y

ClinicalTrialProtocol

- ▶ The protocol associated with a clinical trial;
- ▶ organizes administrative information about the trial
- ▶ such as Organization ID, participants, phase, etc.
- ▶ provides access to the administered Agents.
- ▶ Application: used primarily by CMAP.
- ▶ Related domain objects: Agent, ProtocolAssociation

PROTOCOLOP:clinicaltrial

PROTOCOLOP

Tables Views Synonyms Proc

CLINICALTRIAL: Created: 4/23/2004 10:16:33 AM Last DDL: 4/23/2004 10:16:33 AM

Columns Indexes Constraints Triggers Data Scripts Grants Partitions Sub

Column Name	Col ID	Pk	Data Type	Null?	Default
TITLE	1		VARCHAR2 (500)	N	
PROTOCOLDOCUMENT...	2		VARCHAR2 (50)	Y	
PHASE	3		VARCHAR2 (20)	Y	
PARTICIPATIONTYPE	4		VARCHAR2 (35)	Y	
TREATMENTFLAG	5		VARCHAR2 (5)	Y	
NIHADMINCODE	6		VARCHAR2 (10)	Y	
CURRENTSTATUS	7		VARCHAR2 (40)	Y	
CURRENTSTATUSDATE	8		DATE	Y	
LEADORGANIZATION	9		VARCHAR2 (6)	Y	
LEADORGANIZATIONN...	10		VARCHAR2 (100)	Y	
PINAME	11		VARCHAR2 (85)	Y	
REVIEWERNAME	12		VARCHAR2 (85)	Y	
SOURCE	13		VARCHAR2 (50)	Y	
TYPE	14		VARCHAR2 (50)	Y	
CLINICALTRIALID	15		NUMBER	N	
ISDISABLED	16		NUMBER	Y	0
CTAIND	17		VARCHAR2 (50)	Y	
COMMENTS	18		VARCHAR2 (4000)	Y	

Table

ACCRUAL

ACCRUALCLINICALTRIAL

ACCRUALDEMOGRAPHIC

AGENT

AGENT_CLOSURE

CLINICALTRIAL

CLINICALTRIALAGENT

CLINICALTRIALDOCUMENT

CLINICALTRIALENDPOINT

CLINICALTRIALHISTOLOGY

CLINICALTRIALPROJECTA...

CLINICALTRIALTISSUE

CLINICALTRIALTREATMENT...

CLINICALTRIALTRIALROLE

CONSORTIUM_TYPES

CONTACT

Clone

- ▶ An object used to hold information pertaining to I.M.A.G.E. clones; provides access to sequence information, associated trace files, and the clone's library.
- ▶ Application: imported from the CGAP web site databases.
- ▶ Related domain objects: Sequence, Library, TraceFile, SNP




CABIOOP:clone

	CLONE_ID	CLONE_NAME	V..	I..	LIBRARY_ID	V..	A..	UNIGENE_LIBRARY	C..	C..
▶	2	IMAGE:4734378			17904			6989		
	3	IMAGE:4767316			17904			6989		
	4	GLCFOG07			18970			5601		
	5	IMAGE:1870937			54			1079		
	6	UKCFZH06			19904			6533		
	7	IMAGE:4722638			17904			6989		
	8	IMAGE:2151449			1595	1	Al ⁴	1556		
	9	IMAGE:4712210			17904			6989		
	10	IMAGE:4716802			17904			6989		
	11	IMAGE:4072143			17904			6989		
	12	GLCFOD10			18970			5601		
	13	L17N670205n1-41-A0			29975			12542		
	14	UIH-FG1-bgl-g-02-0-l			27590			11914		
	15	HSI08034			23169			8800		
	17	unknown			22020			8655		
	19	L17N670205n1-15-F1;			29975			12542		
	20	IMAGE:4722596			17904			6989		
	21	IMAGp9981184581_1			54			1079		
	22	GLCCE110			18970			5601		

CMAPOntology

- ▶ An object providing entry to the CMAP gene ontology, which categorizes genes by function; provides access to gene objects corresponding to the ontological term, as well as to ancestor and descendant terms in the ontology tree.
- ▶ Application: defined and used by CMAP applications.
- ▶ Related domain objects: CMAPOntologyRelationship, Gene

CABLOOP:concept

	CONCEPT_ID	CONCEPT_TYPE_ID	CONCEPT_NAME 
	20213	1	BrE-3
	21401	1	C-MYC Expression/Amplification
	20205	1	CD44
	20204	1	CD7 (3A1-antigen)
	20206	1	CEA
	20500	1	CELL CYCLE
	20215	1	CO17-1A
	21501	1	COX-2
	20600	1	CYTOPLASMIC PROTEIN KINASES
	20509	1	Cyclin B Overexpression
	20506	1	Cyclin D Overexpression
	20507	1	Cyclin D1/Plcyclin D1/cdk4
	20508	1	Cyclin E Overexpression
	21510	1	DHFR
	20700	1	DNA INTERACTIVE
	21508	1	DPD
	20310	1	Death receptors DR4
	20320	1	Death receptors DR5
	21002	1	EGFR Overexpression
	21001	1	EGFR expression
	20410	1	Estrogen Receptor
	21201	1	Estrogen Receptor
	21101	1	Estrogen Receptor
	21505	1	FAP

Disease

- ▶ Specifies a disease name and ID; also provides access to: ontological relations to other diseases; clinical trial protocols treating the disease; and specific histologies associated with instances of the disease.
- ▶ Application: used by the CMAP project.
- ▶ Related domain objects: ClinicalTrialProtocol, Histopathology, DiseaseRelationship

ExpressionFeature

- ▶ Associated with a Gene object through the gene's method `getExpressionFeature()`; provides access to the list of organs where the gene is known to be expressed.
- ▶ Application: Expression information for a gene is extracted from the CGAP databases, which are based on the information in Unigene (see discussion of data sources in Chapter 15_The_CaBIO_Data).
- ▶ Related domain objects: Organ, Gene.

ExpressionMeasurement

- ▶ An object representing a structure that is capable of measuring the absolute or relative amount of a given compound.
- ▶ Related domain objects: Gene, Sequence

ExpressionMeasurementArray

- ▶ An array of ExpressionMeasurement objects.
- ▶ Related domain objects: Gene, Sequence

Gene

- ▶ The effective portal to most of the genomic information provided by the caBIO data services;
- ▶ organs, diseases, chromosomes, pathways, sequence data, and expression experiments are among the many objects accessible via a gene.
- ▶ Related domain objects: ExpressionFeature, Organ, Disease, Chromosome, Taxon, Sequence, GeneAlias, GeneHomolog, MapLocation, Protein, SNP, Target, ExpressionMeasurement, Pathway, GoOntology

CABIOOP:gene

GENE_ID	GENE_...	GENE_TITLE	CHROM...	TAXON...	CLUSTER_...	L...	E...	LOCUS_LINK_ID	OMIM_ID
2	NAT2	N-acetyltransferase 2 (arylamine N-acetyltransferase)	0	5	2			10	243400
3	ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	12	5	4			125	103720
4	CEACAM	carcinoembryonic antigen-related cell adhesion molecule 3	2	5	11			1084	
5	CEACAM	carcinoembryonic antigen-related cell adhesion molecule 4	2	5	12			1089	
6	ELA2A	elastase 2A	8	5	21			63036	
7	TGM1	transglutaminase 1 (K polypeptide epidermal type I, protein-glutamine-	4	5	22			7051	190195
8	LTA	lymphotoxin alpha (TNF superfamily, member 1)	9	5	36			4049	153440
10	CEACAM	carcinoembryonic antigen-related cell adhesion molecule 8	2	5	41			1088	
11	PTN	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting	10	5	44			5764	162095
12	PTAFR	platelet-activating factor receptor	8	5	46			5724	173393
13	PIGA	phosphatidylinositol glycan, class A (paroxysmal nocturnal hemoglobi	11	5	51			5277	311770
14	PRPS1	phosphoribosyl pyrophosphate synthetase 1	11	5	56			5631	311850
15	PTPN12	protein tyrosine phosphatase, non-receptor type 12	10	5	62			5782	600079
16	SDHB	succinate dehydrogenase complex, subunit B, iron sulfur (lp)	0	5	64			6390	185470
17	IL1RL1	interleukin 1 receptor-like 1	12	5	66			9173	601203
18	GABPA	GA binding protein transcription factor, alpha subunit 60kDa	10	5	78			2551	600609
19	IL2RG	interleukin 2 receptor, gamma (severe combined immunodeficiency)	11	5	84			3561	308380
20	RBL1	retinoblastoma-like 1 (p107)	15	5	87			5933	116957
21	PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	5	5	96			5366	604959
22	CYP4F2	cytochrome P450, family 4, subfamily F, polypeptide 2	2	5	101			8529	604426
24	HGFAC	HGF activator	1	5	104			3083	604552
25	GRIN1	glutamate receptor, ionotropic, N-methyl D-aspartate 1	6	5	105			2902	138249
26	FGL1	fibrinogen-like 1	0	5	107			2267	605776
27	DPEP1	dipeptidase 1 (renal)	16	5	109			1800	179780
28	KIAA0436	putative L-type neutral amino acid transporter	12	5	110			9581	
29	FGF9	fibroblast growth factor 9 (glia-activating factor)	17	5	111			2254	600921
30	PRDX6	peroxiredoxin 6	8	5	120			9588	602316

GeneAlias

- ▶ An alternative name for a gene; provides descriptive information about the gene (as it is known by this alias), as well as access to the Gene object it refers to.
- ▶ Related Domain Objects: Gene

GeneHomolog

- ▶ Defined only in relation to another Gene object of interest, the functional equivalent of that gene in another taxon (i.e., its ortholog). The GeneHomolog is a specialization of the parent Gene object; in addition to all of the methods provided by the gene interface, the homolog object provides the percent of sequence similarity of the homolog to the related gene of interest.
- ▶ Related domain objects: Gene, ExpressionFeature, Organ, Disease, Chromosome, Taxon, Sequence, GeneAlias, GeneHomolog, MapLocation, Protein, SNP, Target, ExpressionMeasurement, Pathway, GoOntology

CABIOOP:gene_homolog

GENE_ID	RELATED_GENE	PERCENT_ALIGNMENT
2	107445	74.83
3	106799	83.2
6	114526	75.09
7	114085	91.49
8	117163	73.27
9	147957	84.2
11	145552	97.62
12	117236	81.82
14	146296	100
15	133506	84.37
16	137970	92.09
17	147125	67.39
18	108928	96.25
20	136962	90.78
21	143580	69.39
22	108276	79.81
24	110695	81.54
25	144952	99.04
26	144731	82.37
27	109068	72.41
29	108073	99.03
30	126454	89.73
31	107110	91.1
32	108697	77.29
33	135381	98.3
34	106490	66.21
35	140450	81.06
38	107064	78.16
39	100022	86.45

GoOntology

- ▶ An object providing entry to a Gene object's position in the Gene Ontology Consortium's controlled vocabularies, as recorded by LocusLink; provides access to Gene objects corresponding to the ontological term, as well as to ancestor and descendant terms in the ontology tree.
- ▶ Related domain objects: Gene, GoOntologyRelationship

CABIOOP:GoOntology

GO_ID	GO_NAME	HS_GENES	MM_GENES
106	biotin-apoprotein ligase activity	1	1
107	imidazoleglycerol phosphate synthase activity	0	0
108	repairosome	0	0
109	nucleotide excision repair complex	0	0
110	nucleotide excision repair factor 1 complex	0	0
111	nucleotide excision repair factor 2 complex	0	0
112	nucleotide excision repair factor 3 complex	0	0
113	nucleotide excision repair factor 4 complex	0	0
114	G1-specific transcription in mitotic cell cycle	2	0
115	S-specific transcription in mitotic cell cycle	1	0
116	G2-specific transcription in mitotic cell cycle	0	0
117	G2/M-specific transcription in mitotic cell cycle	0	0
118	histone deacetylase complex	14	10
119	mediator complex	13	1
120	RNA polymerase I transcription factor complex	1	0
121	glycerol-1-phosphatase activity	0	0
122	negative regulation of transcription from Pol II promoter	57	42
123	histone acetyltransferase complex	0	3
124	SAGA complex	0	0
125	PCAF complex	0	0
126	transcription factor TFIIIB complex	1	0
127	transcription factor TFIIIC complex	5	0
128	flocculation	0	0
131	incipient bud site	0	0
132	mitotic spindle orientation	0	0
133	polarisome	1	1
134	site of polarized growth (sensu Saccharomyces)	0	0
▶ 135	septin checkpoint	0	0
136	mannosyltransferase complex	0	0
137	Golgi cis cisterna	1	1
138	Golgi trans cisterna	2	0

Histopathology

- ▶ An object representing anatomical changes in a diseased tissue sample associated with an expression experiment; captures the relationship between organ and disease.
- ▶ Application: used by the CMAP project.
- ▶ Related domain objects: Anomaly, Organ, Disease, ExpressionExperiment, ESTExperiment, SAGEExperiment

Library

- ▶ An object representing a CGAP library; provides access to information about: the tissue sample and its method of preparation, the library protocol that was used, the clones contained in the library, and the sequence information derived from the library.
- ▶ Application: Extracted from the CGAP databases.
- ▶ Related domain objects: Clone, Sequence, Tissue, Protocol

MapLocation

- ▶ Associated with a Gene object, the physical map location of the gene.
- ▶ Related domain objects: Chromosome, Gene, Taxon.

Organ

- ▶ A representation of an organ whose name occurs in a controlled vocabulary; provides access to any Histopathology objects for the organ and, because it is “ontolog-able,” provides access to its ancestral and descendant terms in the vocabulary.
- ▶ Related domain objects: Histopathology, OrganRelationship

Pathway

- ▶ An object representation of a molecular/cellular pathway compiled by BioCarta. Pathways are associated with specific Taxon objects, and contain multiple Gene objects, which may be targets for treatment.
- ▶ Related domain objects: Gene, Taxon, TargetTarget.

Protein

- ▶ An object representation of a protein; provides access to the encoding gene via its GenBank ID, the taxon in which this instance of the protein occurs, and references to homologous proteins in other species.
- ▶ Related domain objects: Gene, ProteinHomolog,

ProteinHomolog

- ▶ Defined only in relation to another Protein object of interest, the functional equivalent of that protein in another taxon (i.e., its ortholog). The ProteinHomolog is a specialization of the parent Protein object; in addition to the methods provided by the parent class, the Homolog object provides the percent of sequence similarity of the homolog to the related protein of interest.
- ▶ Related domain objects: Gene, Protein, Taxon.

Protocol

- ▶ An object representation of the protocol used in assembling a clone library.
- ▶ Related domain objects: Library

ProtocolAssociation

- ▶ An association class relating ClinicalTrialProtocols to Diseases.
- ▶ Application: used primarily by the CMAP project.
- ▶ Related domain objects: ConceptSearch, ClinicalTrialProtocol, Disease

Sequence

- ▶ An object representation of a gene sequence; provides access to the clones from which it was derived, the ASCII representation of the residues it contains, and the sequence ID.
- ▶ Related domain objects: Clone, Gene, Protein

SNP

- ▶ A Single Nucleotide Polymorphism; provides access to the clones and trace files from which it was identified, the two most common substitutions at that position, the offset of the SNP in the parent sequence, and a confidence score.
- ▶ Application: identified by the GAI project.
- ▶ Related domain objects: Clone, TraceFile

Target

- ▶ A gene thought to be at the root of a disease etiology and targeted for therapeutic intervention.
- ▶ Application: defined and used by the CMAP project.
- ▶ Related domain objects: Agent, Anomaly, Gene

Taxon

- ▶ An object representing the various names (scientific, common, abbreviated, etc.) for a species associated with a specific Gene, Chromosome, Pathway, Protein, or Tissue object.
- ▶ Related domain objects: Gene, Chromosome, Pathway, Protein, Tissue

Tissue

- ▶ A group of similar cells united to perform a specific function.
- ▶ Related domain objects: Disease, Organ, Protocol, Taxon






**Center
for
Bioinformatics**

Please— ask questions!!



Thank you for using caBIO!!



**Center
for
Bioinformatics**

Home Organization Initiatives Infrastructure Support Download

NCICB :

News


- [NCICB User Applications Manual](#)
- [caBIOperl released](#)
- [caBIG Pilot Structure](#)
- [2003 Jamboree Presentations](#)
- [NCICB Seminars](#)
- [Publications](#)

Quick Links


- [NCICB Support](#)
- [cancer Biomedical Informatics Grid \(caBIG\)](#)
- [Cancer Genome Anatomy Project](#)
- [Cancer Molecular Analysis Project](#)
- [Cancer Models Database](#)
- [Gene Expression Data Portal](#)
- [BIOgopher](#)
- [caBIO Quick Start](#)

The NCI Center for Bioinformatics


The NCI Center for Bioinformatics (NCICB) provides biomedical informatics support and integration capabilities to the cancer research community. We work with both intramural and extramural groups to develop Initiative-Specific Modules. These modules are connected through intelligent interfaces, coordinated through an NCI Core Module and deployed through open source tools and systems. The NCICB also serves as a focal point for cancer research informatics planning worldwide. We work with research organizations, biomedical informatics groups and standards bodies to facilitate the identification and adoption of information exchange standards, thus connecting research information sources wherever they may reside.

**Organization**



Overview of the NCICB's mission, organization, governance, and strategies for executing its charge.

**Supported Research Initiatives**

Access to the WWW sites and related data, tools, and infrastructure developed by the NCICB to support the collection of NCI research initiatives.

**Infrastructure**

[caCORE](#) is the infrastructure backbone supporting data management and application development at NCICB. caCORE includes vocabulary services ([EVS](#)), metadata management ([caDSR](#)), and biomedical data "objects" ([caBIO](#)) implemented in a robust [enterprise software architecture](#). The [caCORE Technical Guide](#) provides detailed documentation of the architecture and [APIs](#). All NCICB-developed caCORE components are [distributed](#) under open-source licenses that support unrestricted usage by both non-profit and commercial entities.



caBIG cancer Biomedical Informatics Grid